Title

Predicting the important enzymes in human breast milk digestion.

Journal

Journal of agricultural and food chemistry, 62(29)

ISSN

0021-8561

Authors

Khaldi, Nora

Vijayakumar, Vaishnavi

Dallas, David C

et al.

Publication Date

2014-07-01

DOI

10.1021/jf405601e

Peer reviewed

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Predicting the Important Enzymes in Human Breast Milk Digestion

Nora Khaldi,*[,†,‡,§] Vaishnavi Vijayakumar,[†,‡] David C. Dallas,[§,⊗] Andrés Guerrero,[○] Saumya Wickramasinghe,[△,⬡] Jennifer T. Smilowitz,[§,⊗] Juan F. Medrano,[⬡] Carlito B. Lebrilla,[○] Denis C. Shields,[†,‡] and J. Bruce German[§,⊗]

[†]Conway Institute of Biomolecular and Biomedical Research, School of Medicine and Medical Sciences and [‡]Complex and Adaptive Systems Laboratory, University College Dublin, Dublin, Republic of Ireland

[§]Department of Food Science and Technology, [⊗]Foods for Health Institute, [○]Department of Chemistry, and [⬡]Department of Animal Science, University of California, Davis, California 95616, United States

[△]Department of Basic Veterinary Sciences, University of Peradeniya, Peradeniya 20400, Sri Lanka

ABSTRACT: Human milk is known to contain several proteases, but little is known about whether these enzymes are active, which proteins they cleave, and their relative contribution to milk protein digestion in vivo. This study analyzed the mass spectrometry-identified protein fragments found in pooled human milk by comparing their cleavage sites with the enzyme specificity patterns of an array of enzymes. The results indicate that several enzymes are actively taking part in the digestion of human milk proteins within the mammary gland, including plasmin and/or trypsin, elastase, cathepsin D, pepsin, chymotrypsin, a glutamyl endopeptidase-like enzyme, and proline endopeptidase. Two proteins were most affected by enzyme hydrolysis: β-casein and polymeric immunoglobulin receptor. In contrast, other highly abundant milk proteins such as α-lactalbumin and lactoferrin appear to have undergone no proteolytic cleavage. A peptide sequence containing a known antimicrobial peptide is released in breast milk by elastase and cathepsin D.

KEYWORDS: *hydrolysate, human milk digestion, milk, nutrition, proteolytic enzymes, bioactive peptide*

INTRODUCTION

Milk is a live secretion that contains numerous complex biomolecules such as proteins, oligosaccharides, and lipids. In addition to these components, previous studies have shown that milk also contains active and inactive forms of hydrolytic enzymes capable of acting upon these biopolymers, including proteolytic enzymes. These independent studies have shown that milk contains plasmin;[1,2] immunoreactive anionic trypsin, most likely present in complex with IgA;[3,4] and cathepsin D.[5]

Although it has been established that proteolytic enzymes are present in milk, little is known about whether these enzymes are active on milk proteins. In addition, if the enzymes were to be active, their contribution to milk protein hydrolysis and their relative contribution compared to one another are unknown.

Many of the studies of milk proteolytic enzymes were carried out on bovine milk, mainly for cheese- or mastitis-related questions; very few were carried out on human milk.[6−9] Most milk proteolytic enzyme investigations have been carried out separately for each enzyme and do not determine the specific effects of each enzyme on milk proteins. This lack of research means that the impact the enzymes have on human milk proteins remains largely unknown. The biological value of these proteolytic enzymes in milk remains ambiguous. Some studies suggest that milk enzymes may support infant growth and nutrition.[10] Some studies suggest that these enzymes may release bioactive peptides that are beneficial for the infant's development (see reviews in refs 11 and 12). But, so far, none of these peptide actions and in vivo catalytic releases have been demonstrated.

Understanding milk composition, and its diverse catalytic activities, notably the proteolytic enzymes, will shed light on mammalian development, evolution, and how to protect neonates. A new generation of analytical and computational tools has made it possible to investigate milk's biopolymers in greater breadth and detail. We previously published a novel mass spectrometry (MS)-based peptide search platform that we used in an iterative searching strategy to identify peptides in minor abundances in human milk.[13] We discovered many peptides, proving that milk proteins can be digested prior to entering the infant's digestive system. Many of these peptides overlapped with known antimicrobial peptides.[13] In this study, we assess the proteolytic activity in human breast milk. We used computational tools to analyze the peptide fragments given their milk protein sequence context. Using these tools, we predicted the proteolytic enzymes that are active in human milk, we examined the relative contribution of each enzyme, and we identified the proteins that have been the most cleaved by the predicted enzymes. mRNA expression was also measured in human milk to determine whether these enzymes are expressed in the mammary gland or whether they enter milk from other sources such as blood.

MATERIALS AND METHODS

Breast Milk Collection. Mature breast milk was collected from two healthy mothers who delivered at term (IRB no. 216198). The milk was kept in storage for up to 1 month. There is no impact on milk proteins and enzymes with storage. In other words, no enzyme activity

Received: December 12, 2013
Revised: March 12, 2014
Accepted: March 12, 2014
Published: March 12, 2014

Table 1. Details of the Enzymes That Take Part in the Digestion of Human Milk Proteins Ranked on the Basis of Their Total Number of Cleavages[a]

enzyme[b]	N-terminal cleavage count	C-terminal cleavage count	total cleavage	unique cleavage	no. of expected cleavages within the peptide	no. of proteins cleaved	odds ratio	std error	gene expression given in RPKM
plasmin	120	200	320	0	371	15	4.11	1.08	0
trypsin 1[c]	120	199	319	0	352	14	4.33	1.09	0 (5.81 PRKM for trypsin domain containing protease TYSND1)
cathepsin D	68	62	130	0	846	24	0.61	1.1	365.07
chymotrypsin low 1	68	34	102	0	449	11	0.93	1.12	0
elastase	51	39	90	0	563	13	0.64	1.12	0
pepsin 1 (pH 1.3)	31	47	78	0	479	10	0.66	1.13	0.1
glutamyl endopeptidase[c]	55	21	76	0	930	5	0.31	1.13	0 (75.97 PRKM for proteasome subunit β type-6, responsible for the peptidyl glutamyl-like activity, PSMB6)
pepsin 1 (pH >2)	30	38	68	0	315	8	0.89	1.15	0.1
proline endopeptidase	14	31	45	0	598	18	0.29	1.17	5.52
chymotrypsin low 4	**3**	**12**	**15**	**0**	**102**	**4**	**0.60**	**1.32**	**0**
chymotrypsin low 3	**6**	**6**	**12**	**0**	**79**	**6**	**0.62**	**1.36**	**0**
chymotrypsin low 2	**2**	**0**	**2**	**0**	**0**	**1**	**20.65**	**4.71**	**0**

[a]The corresponding gene expression levels in RPKM are provided for each enzyme. [b]Enzymes in bold represent ones that are not expressed at day 60 of lactation, but low expression of these genes has been detected at earlier stages of lactation (15 and 30 days). [c]Enzymes that do not show expression in milk, but expression data shows the presence of other enzymes that have a similar activity and are expressed.

occurs when samples are frozen. The stage of lactation of the first mother was 3 months and that of the second was 4 months. The mothers providing this milk were instructed to cleanse each breast with water and then use an electronic pump to collect a pool of milk from both breasts. The samples were then stored in the subjects' freezers and delivered to the laboratory on ice where they were stored at −80C.

Sample Preparation. Peptides were extracted from human milk according to the method of Dallas et al.[13] Briefly, 200 μL of each milk sample was thawed on ice and combined. To remove milk fat globules, the milk was centrifuged at 3000*g* for 10 min and the skim infranate was extracted. Centrifugation was repeated on the skim infranate to remove any remaining visible lipid layer. Proteins were then precipitated with the addition of 400 μL of 200 g/L trichloroacetic acid. Samples were vortexed briefly and then centrifuged at 3000*g* for 10 min, and the peptide-containing supernatant was collected, leaving the precipitated protein. This precipitation was repeated for a total of three times. Trichloroacetic acid, salts, oligosaccharides, and lactose were then removed from the peptides by C18 reverse-phase preparative chromatography. Contaminants were eluted with water, and peptides were then eluted with 80% acetonitrile (ACN)/0.1% trifluoroacetic acid (v/v). The peptide solution was then dried down in a vacuum centrifuge at 37 °C. After drying, the sample was rehydrated in 40 μL of nanopure water for MS analysis.

MS Analysis of Human Breast Milk Peptides. Peptides were analyzed via nanoliquid chromatography chip quadrupole time-of-flight tandem MS (Agilent, Santa Clara, CA, USA). Two microliters of sample was injected for each run onto the C18 reverse-phase nanochip. The nanopump flow was 0.3 μL/min, and the capillary pump flow rate was 3 μL/min. Peptides were eluted with the following gradient of solvent A (3% ACN/0.1% formic acid (FA) (v/v)) and solvent B (90% ACN/0.1% FA (v/v)): 0−8% B from 0 to 5 min, 8−26.5% B from 5 to 24 min, 26.5−100% B from 24 to 48 min, followed by 100% B for 2 min and 100% A for 10 min (to re-equilibrate the column). The instrument was run in positive ionization mode. Data collection thresholds were set at 200 ion counts or 0.01% relative intensity for MS spectra and at 5 ion counts or 0.01% relative intensity for MS/MS. Data were collected in centroid mode. The drying gas was 350 °C, and flow rate was 3 L/min. The required chip voltage for consistent spray varied from 1850 to 1920 V. Automated precursor selection based on abundance was employed to select peaks for tandem fragmentation with an exclusion list consisting of all peptides identified in previous analyses in this study. The acquisition rate employed was 3 spectra/s for both MS and MS/MS modes. The isolation width for tandem analysis was 1.3 m/z. The collision energy was set by the formula (slope) × $(m/z)/100$ + offset, with slope = 3.6 and offset = −4.8. Five tandem spectra were collected after each MS spectrum, with active exclusion after 5 MS/MS for 0.15 min. Precursor ions were selected only if they had at least 1000 ion counts or 0.01% of the relative intensity of the spectra. Mass calibration was performed during data acquisition based on an infused calibrant ion with a mass of 922.009789 Da.

Agilent Mass Hunter Qualitative Analysis Software (Santa Clara, CA, USA) was used to analyze the data obtained. Molecules identified in the spectral analysis were grouped into compounds by the Find by Molecular Feature algorithm, which groups together molecules across charge state and charge carrier. All tandem MS from each data file were exported as Mascot Generic Files (.mgf) with a peptide isotope model and a maximum charge state of +9.

Peptide identification was accomplished using both the MS-GFDB (via a command-line interface) and X!Tandem (using the downloadable graphical user interface). The human milk library used in both searches was constructed on the basis of a query to the Uniprot database. The query returned only proteins from *Homo sapiens* and at least one of the following: "tissue specificity" keyword "milk" or "mammary", "tissue" keyword "milk" or "mammary", or gene ontology "lactation". This query returned a list of 1472 proteins. These were exported to FASTA file format. For MS-GFDB, peptides were accepted if p values were ≤0.01 corresponding to confidence levels of 99%. No p values exist in X!Tandem, so a closely related statistic, e value, was used for the X!Tandem search. The e value thresholds selected were again 0.01, reflecting 99% confidence. In both programs, masses were allowed 20 ppm error. No complete (required) modifications were included but up to four potential modifications were allowed on each peptide. Potential modifications allowed were phosphorylation of serine, threonine, or tyrosine and oxidation of methionine. A nonspecific cleavage ([X]|[X]) (where "X" is any amino acid) was used to search against the protein sequences. For MS-GFDB, the fragmentation method selected in the search was collision-induced dissociation and the instrument selected was time-of-flight. For X!Tandem, there was no option for fragmentation type and instrument selection. Because the instrument did not always select the monoisotopic ion for tandem fragmentation, isotope errors were allowed (allowing up to one C13). No model refinement was employed in X!Tandem.

Enzyme Prediction. The web-based software EnzymePredictor[14] was employed to evaluate and predict which enzymes most likely contributed to cleavage of human breast milk proteins (http://bioware.ucd.ie/~enzpred/Enzpred.php). Enzymes were classified on the basis of their total number of performed cleavages, and they were evaluated on the basis of their odds ratio (OR; see Table 1), which is an indicator of their degree of participation in the hydrolysis of the proteins. The OR values indicate that certain enzymes are over-represented and others under-represented. The following information was also collected from EnzymePredictor: number of times an enzyme could have cleaved within the current peptides, total number of proteins cleaved by each enzyme, total number of cleavages performed by an enzyme on the C- and N-termini.

Expression Profiling of Human Milk. *Fresh Milk Sample Collection.* Fresh milk samples were obtained from three healthy females on days 4, 15, 30, and 60 postpartum who gave birth to a term infant (>37 weeks of gestation). In the early morning period, the donor manually pumped one breast until emptied into a collection bag, which was immediately delivered on cold-packs to the laboratory for processing. The Institutional Review Board of University of California, Davis, approved the project.

RNA Extraction for Gene Expression Studies. Somatic cells were pelleted by adding 50 μL of 0.5 M ethylenediaminetetraacetic acid to 20 mL of fresh milk and centrifuged at 2000*g* at 4 °C for 10 min.[15] The pellet of cells was washed with 10 mL of phosphate-buffered saline at pH 7.2 and 10 μL of 0.5 M ethylenediaminetetraacetic acid (final concentration = 0.5 mM) and filtered through sterile cheesecloth to remove any debris. The cells were then centrifuged again at 2000*g* at 4 °C for 10 min. The supernatant was decanted, and RNA was extracted from the milk somatic cell pellet using Trizol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. RNA was quantified by an ND-1000 spectrophotometer (Fisher Thermo, Wilmington, MA, USA), and the quality and integrity were assessed by the spectrophotometer 260/280 ratio, gel electrophoresis, and capillary electrophoresis with an Experion Bioanalyzer (Bio-Rad, Hercules, CA, USA).

RNA Sequencing and Data Analysis. Gene expression analysis was conducted on fresh milk samples collected on days 4, 15, 30, and 60 postpartum by RNA sequencing (RNA-Seq). Messenger RNA was isolated and purified using an RNA-Seq sample preparation kit (Illumina, San Diego, CA, USA). The fragments were purified and sequenced at the UC Davis Genome Center DNA Technologies Core Facility using the Illumina Genome Analyzer (GAII) and Illumina HiSeq 2000. Sequence reads were assembled and analyzed in RNA-Seq. Expression analysis was performed with the CLC Genomics Workbench 6.0 (CLC Bio, Aarhus, Denmark). Human genome, GRCh37.69 (ftp://ftp.ensembl.org/pub/release-69/genbank/homo_sapiens/), was utilized as the reference genome for the assembly. Data were normalized by calculating the "reads per kilobase per million mapped reads" (RPKM)[16] for each gene and annotated with ENSEMBL human genome assembly (55203 unique genes).

RESULTS AND DISCUSSION

This project aimed to determine the enzymes responsible for the cleavage patterns identified in milk, examine each enzyme's contribution to the hydrolyses of human milk proteins, and clarify whether these enzymes are expressed in milk or enter milk from other sources.

A novel MS-based peptide search platform using an iterative searching strategy detected a variety of peptides in minor abundances in pooled human breast milk samples.[13] The peptides extracted and identified for the present study were analyzed using EnzymePredictor.[14] The results of this analysis are presented in Table 1. The enzymes are ordered on the basis of their total number of cleavages, from those that have cleaved extensively to those that are predicted to have cleaved one residue. As discussed in the interpretation of EnzymePredictor,[14] the enzymes with a combined high number of total cleavages performed and a high odds ratio are the enzymes that are most likely active in the milk.

Expression profiles from samples of human breast milk were established as the means to investigate if the predicted enzymes originate from the mammary epithelial cells or they enter milk from other sources, such as blood. The samples that were obtained for analysis for peptide detection via MS were pooled from mature milk samples from two mothers, both 3 months postpartum, whereas that of expression data was of 2 month postpartum milk.

Major Cleavage of Proteins in Human Milk Is at Trypsin and Plasmin Target Sites. The natural cleavages of milk proteins showed a strong enrichment for cleavage after K or R, both consistent with plasmin and trypsin cleavage (Tables 1 and 4).

The presence of plasmin, an enzyme that plays a major role in the proteolytic breakdown of blood clots, has been previously reported in milk,[1] but its contribution to human milk proteolytic hydrolysis is unknown. This study found that plasmin potentially cleaved 320 peptides (Figure 1; Table 1)

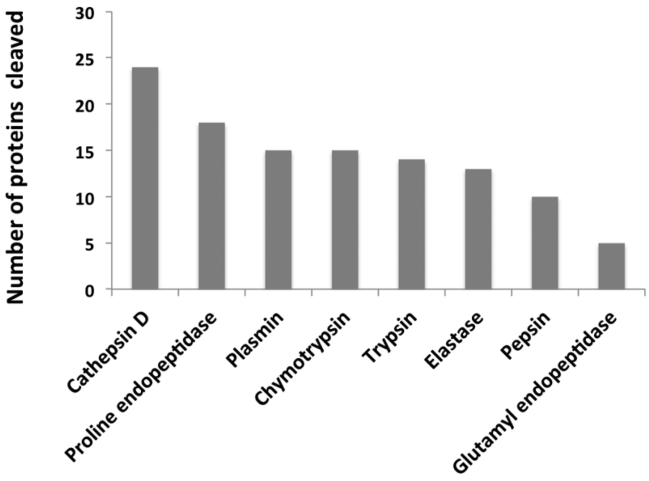

Figure 1. Representation of the total number of cleaved milk proteins per predicted enzyme. Each enzyme name on the *X*-axis is plotted against the total number of proteins it cleaves on the *Y*-axis.

derived from 15 milk proteins (Table 2). Anionic trypsin's presence in human milk was initially identified by Monti et al. and Borulf et al.[3,4] Borulf et al. showed that trypsin's presence in milk is likely in complex with IgA. Whether this enzyme contributes to the hydrolysis of human milk proteins remains unknown.

The two proteins most affected by potential plasmin or trypsin digestion were polymeric immunoglobulin receptor (32 cleavages of peptides) and β-casein (89). Other hydrolyzed proteins included α_{S1}-casein (11), osteopontin (18), butyrophilin subfamily 1 member A1 (15), and κ-casein (7). These analyses found that plasmin cleaves an androgen receptor that trypsin failed to cleave (Q9UN21; Tables 2 and 3). Butyrophilin subfamily 1 member A1 is part of the milk fat globule membrane. Thus, the low number of peptides detected from this protein might be due to the centrifugation for the removal of the milk fat globules. Trypsin can typically not cleave if the K or R is followed by a P, which makes the androgen receptor better fit the plasmin pattern (Table 4). This result would suggest that the cleavage tends to follow the specificity of plasmin more consistently than that of trypsin.

dx.doi.org/10.1021/jf405601e | J. Agric. Food Chem. 2014, 62, 7225−7232

Table 2. Human Milk Proteins Found To Have Been Digested by Plasmin[a]

protein name	Uniprot ID and access name	total no. of cleavages performed
β-casein	P05814 (CASB_HUMAN)	89
polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	32
osteopontin	P10451 (OSTP_HUMAN)	18
butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	15
α_{S1}-casein	P47710 (CASA1_HUMAN)	11
κ-casein	P07498 (CASK_HUMAN)	7
mucin-1	P15941 (MUC1_HUMAN)	3
parathyroid hormone-related protein	P12272 (PTHR_HUMAN)	2
bile salt-activated lipase	P19835 (CEL_HUMAN)	2
androgen receptor	Q9UN21 (Q9UN21_HUMAN)	1
protein diaphanous homologue 1	O60610 (DIAP1_HUMAN)	1
complement C4-A	P0C0L4 (CO4A_HUMAN)	1
La-related protein 1	Q6PKG0 (LARP1_HUMAN)	1
NMDA receptor-regulated protein 2	Q659A1 (NARG2_HUMAN)	1
dedicator of cytokinesis protein 1	Q14185 DOCK1_HUMAN	1

[a]The proteins are ordered on the basis of the number of cleavages that plasmin has performed on each. Here we considered only the number of peptides that are unique; if a peptide was not unique, we considered only one copy of the peptide. Proteins exclusively found in milk are presented in bold.

Table 3. Human Milk Proteins Found To Have Been Digested by Trypsin[a]

protein name	Uniprot ID and access name	total no. of cleavages performed
β-casein	P05814 (CASB_HUMAN)	89
polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	32
osteopontin	P10451 (OSTP_HUMAN)	18
butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	15
α_{S1}-casein	P47710 (CASA1_HUMAN)	11
κ-casein	P07498 (CASK_HUMAN)	7
mucin-1	P15941 (MUC1_HUMAN)	3
parathyroid hormone-related protein	P12272 (PTHR_HUMAN)	2
bile salt-activated lipase	P19835 (CEL_HUMAN)	2
La-related protein 1	Q6PKG0 (LARP1_HUMAN)	1
protein diaphanous homologue 1	O60610 (DIAP1_HUMAN)	1
NMDA receptor-regulated protein 2	Q659A1 (NARG2_HUMAN)	1
complement C4-A	P0C0L4 (CO4A_HUMAN)	1
dedicator of cytokinesis protein 1	Q14185 (DOCK1_HUMAN)	1

[a]The proteins are ordered on the basis of the number of cleavages that trypsin has performed on each. Here we considered only the number of peptides that are unique; if a peptide was not unique, we considered only one copy of the peptide. Proteins exclusively found in milk are presented in bold.

Table 4. Enzyme Specificity for Plasmin and Trypsin

	cleavage pattern						
enzyme	P4	P3	P2	P1	P1′	P2′	ref
trypsin	–	–	W	K	P	–	22
	–	–	M	R	P	–	22
plasmin	–	–	–	K or R	–	–	23

However, there are too few cleavages representing K or R followed by P (only one in our case) to disambiguate between plasmin and trypsin activities in the milk.

Enzymes Cleaving Hydrophobic Target Sites: Elastase, Cathepsin D, Pepsin, and Chymotrypsin. Whereas the active form of cathepsin D is found in bovine milk,[5] the major form of this enzyme in milk is the inactive zymogen, procathepsin D.[17] In this study, cathepsin D was found to be active in human milk (Table 1), cleaving 130 times. Cathepsin D hydrolyzed the highest number of milk proteins (24 milk proteins) compared to the other enzymes present in milk (Tables 1 and 5; Figure 1). Although the total number of times cathepsin D performed a cleavage is high (130 total cleavages), most of these are located in two proteins: β-casein with 78 cleavages and polymeric immunoglobulin receptor with 21 cleavages. Interestingly, both of these proteins have also been the most digested ones with all five top predicted enzymes.

A high number of potential cleavage sites was predicted for cathepsin D (846; Table 1), which were not observed in the peptide results, with an odds ratio below 1. This failure to produce predicted peptides could be due to the positioning of a subpopulation of these residues in regions difficult for the enzyme to access due to structural constraints. This structural inhibition is more likely for cathepsin D sites, because many potential cleavage sites are strongly hydrophobic, and hydrophobic regions are usually buried within the structured regions of proteins.

Elastase is another enzyme predicted to be active in human milk. Elastase is known to be a major proteinase in polymorphonuclear neutrophils (PMNs), which are phagocytic cells that destroy infectious agents in humans.[18] Elastase activity has been detected in PMNs recovered from milk during experimentally induced mastitis.[19] There is substantial overlap between the cleavage sites of elastase and cathepsin D, and again there are a large number of sites (563) interior to the peptides that were not cleaved. Elastase potentially performed 90 total cleavages of milk peptides (Table 6) over 13 proteins. The proteins that elastase cleaved the most are β-casein (41 cleavages of unique peptides) and polymeric immunoglobulin receptor (25 cleavages of unique peptides). No traces of elastase activity were observed on κ-casein (Table 6), and only one cleavage was observed for cathepsin D (Table 5). This reflects the similar, but not identical, cleavage patterns between elastase and cathepsin D. To resolve the specificity, the number of sites that were cleaved in common between both was determined (60 cleavages of unique peptides) or asa were those specific to only one (69 for cathepsin D of unique peptides, and 28 for elastase of unique peptides). This pattern suggests a role for both enzymes in human milk digestion prior to infant consumption.

Two other enzymes with hydrophobic targets, pepsin and chymotrypsin, are consistent with a number of digestion sites in addition to those digestible by cathepsin D and elastase. Only

Table 5. Human Milk Proteins Found To Have Been Digested by Cathepsin D[a]

protein name	Uniprot ID and access name	total no. of cleavages performed
β-casein	P05814 (CASB_HUMAN)	78
polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	21
perilipin-2	Q99541 (PLIN2_HUMAN)	3
osteopontin	P10451 (OSTP_HUMAN)	2
α_{S1}-casein	P47710 (CASA1_HUMAN)	2
deubiquitinating protein VCIP135	Q96JH7 (VCIP1_HUMAN)	2
butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	2
receptor-type tyrosine-protein phosphatase	Q15262 (PTPRK_HUMAN)	2
macrophage mannose receptor 1	P22897 (MRC1_HUMAN)	2
protein CASC3	O15234 (CASC3_HUMAN)	1
flavin containing monooxygenase 5, isoform CRA_c	Q9HA79 (Q9HA79_HUMAN)	1
Abl interactor 1	Q8IZP0 (ABI1_HUMAN)	1
misshapen-like kinase 1	Q8N4C8 (MINK1_HUMAN)	1
La-related protein 1	Q6PKG0 (LARP1_HUMAN)	1
receptor-type tyrosine-protein phosphatase α	P18433 (PTPRA_HUMAN)	1
ubiquitin carboxyl-terminal hydrolase 51	Q70EK9 (UBP51_HUMAN)	1
protein diaphanous homologue 1	O60610 (DIAP1_HUMAN)	1
neural Wiskott−Aldrich syndrome protein	O00401 (WASL_HUMAN)	1
κ-casein	P07498 (CASK_HUMAN)	1
insulin receptor substrate 1	P35568 (IRS1_HUMAN)	1
transcription factor 7-like 2	Q9NQB0 (TF7L2_HUMAN)	1
γ-glutamyltransferase 6	Q6P531 (GGT6_HUMAN)	1
PH domain leucine-rich repeat-containing protein phosphatase 1	O60346 (PHLP1_HUMAN)	1
dedicator of cytokinesis protein 1	Q14185 (DOCK1_HUMAN)	1

[a]The proteins are ordered on the basis of the number of cleavages that trypsin has performed on each. Here we considered only the number of peptides that are unique; if a peptide was not unique, we considered only one copy of the peptide. Proteins exclusively found in milk are presented in bold.

Table 6. Human Milk Proteins Found To Have Been Digested by Elastase[a]

protein name	Uniprot ID and access name	total no. of cleavages performed
β-casein	P05814 (CASB_HUMAN)	41
polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	25
butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	8
αS1-casein	P47710 (CASA1_HUMAN)	3
perilipin-2	Q99541 (PLIN2_HUMAN)	3
protein CASC3	O15234 (CASC3_HUMAN)	1
androgen receptor	Q9UN21 (Q9UN21_HUMAN)	1
parathyroid hormone-related protein	P12272 (PTHR_HUMAN)	1
osteopontin	P10451 (OSTP_HUMAN)	1
heat shock protein β-1	P04792 (HSPB1_HUMAN)	1
receptor-type tyrosine-protein phosphatase K	Q15262 (PTPRK_HUMAN)	1
receptor-type tyrosine-protein phosphatase α	P18433 (PTPRA_HUMAN)	1
γ-glutamyltransferase 6	Q6P531 (GGT6_HUMAN)	1

[a]The proteins are ordered on the basis of the number of cleavages that trypsin has performed on each. Here we considered only the number of peptides that are unique; if a peptide was not unique, we considered only one copy of the peptide. Proteins exclusively found in milk are presented in bold.

21 of the 75 total peptide terminal cleavage sites of unique peptides that are predicted to be performed by chymotrypsin are common cleavages with elastase and cathepsin D. For pepsin, only 38 of the 72 cleavages sites predicted from the unique peptides are shared with elastase and cathepsin D. Furthermore, only 16 peptide terminal cleavage sites are shared between both pepsin and chymotryspin. These findings indicate the first support that pepsin- and chymotrypsin-like activities are present in human breast milk (Table 1).

Examination of Gene and Protein Expression Profile of Proteases in Human Milk. Previous work has suggested that plasmin is found in milk but that it originates from blood.[20] This study tested if the enzymes responsible for the observed peptide fragments are produced in the mammary gland or migrate into milk from other origins. Gene expression analysis was carried out for human milk according to the procedure used for bovine milk[15] (Table 1). The expression was analyzed for days 4, 15, 30, and 60 of lactation. For consistency purposes, day 60 gene expression was determined; however, we also took into account the expression levels for the other days of lactation to build Table 1, which shows the possibility of expression of an enzyme at different stages of lactation. The results show that very few of the proteolytic enzymes described in Table 1 are expressed in the mammary gland at these time points (Table 1). In agreement with previous literature, no gene expression of plasmin was detected in the milks. Moreover, no expression was observed for the intestinal enzyme trypsin.

Cathepsin D was highly expressed (365 RPKM, Table 1). Elastase, however, was not expressed in the mammary epithelial cells (Table 1). As mentioned above, the presence of active elastase in milk may be explained by its presence in PMNs. We would expect the vast majority of PMN cells to be pelleted by the centrifugation and, thus, would not participate in any degradation of the milk proteins after centrifugation (see Materials and Methods). However, PMN can secrete elastase

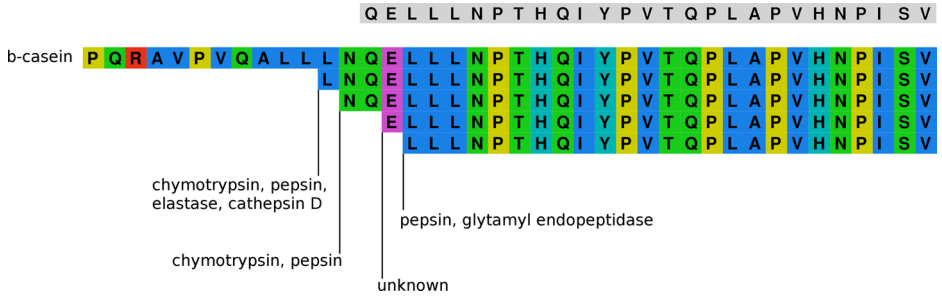

Figure 2. Illustration of antimicrobial-like peptides released in vivo from β-casein in human milk. The peptide QELLLNPTHQIYPVTQPLAPV-HNPISV shown at the top of the figure in gray has been discovered to be an antimicrobial peptide.[21] This peptide is found at the C-terminus of β-casein (positions 199−225). Four overlapping peptides found to be released in vivo in human milk are represented under the β-casein sequence. The in vivo cleavage positions are indicated with gray lines. The enzymes predicted to be responsible for these cleavages are presented beside the gray lines.

while still within the mammary gland, and the predicted elastase activity may derive from it.

Other Predicted Proteases That Were Also Expressed in Human Milk. Most of the other predicted enzymes have a low number of total cleavages, making it difficult to fully support their presence in milk (Table 1; total cleavages $\leq$ 44). Interestingly, some of these enzyme activities predicted in milk seem to be expressed in matching proteases. On the basis of a relatively specific cleavage site not shared by other proteases in these analyses (cleaves after E; with 76 cleavages in total), we predicted a glutamyl endopeptidase-like activity. Transcripts for glutamyl endopeptidase were, however, not detected in mammary epithelial cells. A glutamyl endopeptidase-like protease (proteasome subunit β type-3, PSMB3) is, however, highly expressed (75.9 RPKM; Table 1). This result may explain the high number of cleavage sites (930) predicted to be due to a glutamyl endopeptidase-like enzyme. Indeed, this glutamyl endopeptidase-like protease may have a more specific cleavage pattern explaining the many residues containing glutamic acid (E) that it did not cleave, as this cannot be accounted for by a structural bias, because the charged glutamic acids are not found buried in the core of the proteins. Although the gene expression of a glutamyl endopeptidase (called "a disintegrin and metalloproteinase with thrombospondin motifs 4" (ADAMTS4) was detected, its potential cleavage site, E'[AS], is cleaved only eight times in the data set. Thus, there is no indication for this enzyme playing a role as the major glutamyl endopeptidase activity in milk.

Proline-endopeptidase was found to be expressed in milk (5.52 RPKM; Table 1). Proline-endopeptidase is responsible for the cleavage of 45 residues covering a total of 18 milk proteins (Table 1; Figure 1).

In Vivo Release of Antimicrobial Peptides in Human Milk. Milk proteins may carry encrypted functional peptide sequences that, when released by enzymes from the intact protein, help in the protection and development of the neonate.[12] However, without in vivo results, these concepts remain unproven. This study identified four peptides that overlap (two extra or fewer amino acids on the N-terminus of the peptide) with a known antimicrobial peptide that has been reported previously in the literature.[21] This peptide is present at the C-terminus of β-casein (Figure 2). These four overlapping peptides are naturally released in human milk via the action of several enzymes, including cathepsin D and elastase (Figure 2). The overlapping peptides most likely carry the antibacterial activity, as the amino acid additions compared to the literature-defined sequence are unlikely to abolish the antimicrobial activity of these sequences.

Uneven Distribution of Enzyme Activity in Human Milk. To measure the enzyme activity, we considered the unique cleavages of each (i.e., if an enzyme cleaves out the same peptide from the parent protein multiple times, we will consider this to be only one unique cleavage). This measure highlights more the range of cleavages of an enzyme and makes it possible to compare it with other enzyme ranges rather than comparing its ability to cleave many times the same peptides. Consequently, this measure does not take the abundance of the proteins into account. Using this approach, we find that two proteins, β-casein and polymeric immunoglobulin receptor, show the highest susceptibility to the milk enzyme activity, as shown by the large number of fragments found from these two proteins. β-Casein is a milk-specific protein expressed during lactation. Interestingly, the other two milk-specific proteins, α_{S1}-casein and κ-casein, show little digestion susceptibly within the mammary gland. The possibility of protein structural disorder being the source of variability in degradation susceptibility among these proteins is not supported by these results, as κ-casein and α_{S1}-casein are also highly disordered proteins. The K and R residues are those that show the most susceptibility for the cleavages observed in human milk proteins. However, no enrichment of these amino acids was found in β-casein versus α_{S1}-casein or κ-casein (14 K and R sites in β-casein; 16 in α_{S1}-casein; and 13 in κ-casein). In fact, there are even fewer K and R in β-casein over the sequence length compared to α_{S1}-casein or κ-casein, arguing that this is not the driver behind β-casein susceptibility, nor is it for polymeric immunoglobulin receptor (6% in β-casein, 9% in α_{S1}-casein, 8% κ-casein, and 10% in polymeric immunoglobulin receptor). On the other hand, post-translational modifications may also explain these variations in susceptibility of proteins to degradation. Indeed, large modifications on proteins may prevent enzymes from reaching their target site. However, investigating the potential modification sites using Uniprot shows that all of the casein proteins are slightly/equivalently modified (in terms of numbers of glycosylations and phosphorylations).

The other major milk-specific proteins that have not been affected are α-lactalbumin and lactoferrin. No peptides from these proteins were detected despite the fact that plasmin, trypsin, cathepsin D, and elastase were all predicted to cleave both of these proteins. It is perhaps the particular structure and post-translational modifications of α-lactalbumin and lactoferrin that prevented the enzymes from cleaving them.

Although cathepsin D is highly expressed in milk (Table 1), it does not show a high effectiveness in cleaving milk proteins. This lower cleavage rate may be because cathepsin D is most effective at a more acidic pH (pH 5 for cathepsin D), which is different from the more neutral milk pH.

Independent studies have shown the existence and activity of some enzymes in milk, but the extent to which these enzymes are active relative to each other remained unknown. Likewise, whether or not these enzymes are created in the mammary epithelial cells was unclear. The enzymes that are active in human milk and responsible for protein digestion prior to entry into the infant's digestive system were determined. Using a combination of bioinformatics, peptidomics, transcriptomics, and literature results, we showed that milk begins to be digested even before infant consumption and that this digestion is mainly carried out by four proteolytic enzymes: plasmin, a trypsin-like enzyme, elastase, and cathepsin D (Tables 1−3, 5, and 6). Among these four enzymes, only cathepsin D and elastase are expressed in milk, with elastase showing only weak expression at days 15 and 30 and no expression after that (Table 1). Activity and expression of chymotrypsin and pepsin-like activity were also seen (Table 1). This finding is surprising, given that these enzymes have never been reported in human breast milk. Other enzymes that this analysis predicted from activity and expression are glutamyl endopeptidase and proline endopeptidase (Table 1).

Interestingly, plasmin, trypsin, elastase, and cathepsin D cleave the casein proteins (α_{S1}-, β-, and κ-casein) to various extents. Our novel MS-based peptide search platform using an iterative searching strategy to detect peptides in minor abundances did not detect peptides resulting from α-lactalbumin. Plasmin, a trypsin-like enzyme, elastase, and cathepsin are predicted to cleave this protein, yet none of the predicted peptides could be detected. The mechanism causing this protein selectivity of digestion remains unknown.

We know very little about proteolytic enzyme activity in human milk, which makes the problem of humanizing milk formulas even harder. The work we have carried out here sheds light on the different enzymes we find in human milk and the relative contribution of each of them, in terms of the release of unique peptides, to cleaving milk proteins. The example illustrated in Figure 2 of a known antimicrobial activity being released illustrates how enzyme activity in human milk may be playing a much greater role than previously anticipated, and fully clarifying this role is key to understanding infant needs.

AUTHOR INFORMATION

Corresponding Author

*(N.K.) E-mail: nora.khaldi@ucd.ie. Phone: 00353-1-7165335. Mail: UCD Complex and Adaptive Systems Laboratory, University College Dublin, Dublin 4, Ireland.

Funding

This work was funded by the Irish Research Council for Science, Engineering and Technology, cofunded by Marie Curie Actions under FP7; Enterprise Ireland grant to Food Health Ireland; SFI grant 08/IN.1/B1864; University of California Discovery Program (05GEB01NHB), the National Institutes of Health (HD059127 and HD061923) California Dairy Research Foundation; U.S. Department of Agriculture, National Institute for Food and Agriculture postdoctoral fellowship; National Science Foundation Graduate Research Fellowship Program; and National Institutes of Health Training Program in Biomolecular Technology (2-T3-GM08799).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. Alan Kelly and Kevin Rue for helpful discussions.

ABBREVIATIONS USED

MS, mass spectrometry; FA, formic acid; ACN, acetonitrile; OD, odds ratio; PMN, polymorphonuclear neutrophils; RPKM, reads per kilobase per million mapped reads

REFERENCES

(1) Heegaard, C. W.; Larsen, L. B.; Rasmussen, L. K.; Højberg, K. E.; Petersen, T. E.; Andreasen, P. A. Plasminogen activation system in human milk. *J. Pediatr. Gastroenterol. Nutr.* **1997**, *25*, 159−166.

(2) Nielsen, S. S. Plasmin system and microbial proteases in milk: characteristics, roles, and relationship. *J. Agric. Food Chem.* **2002**, *50*, 6628−6634.

(3) Borulf, S.; Lindberg, T.; Mansson, M. Immunoreactive anionic trypsin and anionic elastase in human-milk. *Acta Paediatr.* **1987**, *76*, 11−15.

(4) Monti, J.; Mermoud, A.; Jolles, P. Trypsin in human milk. *Experientia* **1986**, *42*, 39−41.

(5) Kaminogawa, S.; Yamauchi, K. Acid protease of bovine milk. *Agric. Biol. Chem.* **1972**, *36*, 2351−2356.

(6) Armaforte, E.; Curran, E.; Huppertz, T.; Ryan, C. A.; Caboni, M. F. O.; Connor, P. M.; Ross, R. P.; Hirtz, C.; Sommerer, N.; Chevalier, F. Proteins and proteolysis in pre-term and term human milk and possible implications for infant formulae. *Int. Dairy J.* **2010**, *20*, 715−723.

(7) Ferranti, P.; Traisci, M. V.; Picariello, G.; Nasi, A.; Boschi, V.; Siervo, M.; Falconi, C.; Chianese, L.; Addeo, F. Casein proteolysis in human milk: tracing the pattern of casein breakdown and the formation of potential bioactive peptides. *J. Dairy Res.* **2004**, *71*, 74−87.

(8) Borulf, S.; Lindberg, T.; Mansson, M. Immunoreactive anionic trypsin and anionic elastase in human milk. *Acta Paediatr. Scand.* **1987**, *76*, 11−15.

(9) Monti, J. C.; Mermoud, A. F.; Jolles, P. Trypsin in human milk. *Experientia* **1986**, *42*, 39−41.

(10) Shahani, K.; Kwan, A.; Friend, B. A. Role and significance of enzymes in human milk. *Am. J. Clin. Nutr.* **1980**, *33*, 1861−1868.

(11) Korhonen, H.; Pihlanto, A. Bioactive peptides: production and functionality. *Int. Dairy J.* **2006**, *16*, 945−960.

(12) Dallas, D. C.; Underwood, M. A.; Zivkovic, A. M.; German, J. B. Digestion of protein in premature and term infants. *J. Nutr. Disorders Ther.* **2012**, *2*, 1−9.

(13) Dallas, D. C.; Guerrero, A.; Khaldi, N.; Castillo, P. A.; Martin, W. F.; Smilowitz, J. T.; Bevins, C. L.; Barile, D.; German, J. B.; Lebrilla, C. B. Extensive in vivo human milk peptidomics reveals specific proteolysis yielding protective antimicrobial peptides. *J. Proteome Res.* **2013**, *12*, 2295−2304.

(14) Vijayakumar, V.; Guerrero, A.; Davey, N.; Lebrilla, C. B.; Shields, D. C.; Khaldi, N. EnzymePredictor: a tool for predicting and visualizing enzymatic cleavages of digested proteins. *J. Proteome Res.* **2012**, *11*, 6056−6065.

(15) Wickramasinghe, S.; Rincon, G.; Islas-Trejo, A.; Medrano, J. F. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* **2012**, *13*, 45−58.

(16) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621−628.

(17) Larsen, L. B.; Petersen, T. E. Identification of five molecular forms of cathepsin D in bovine milk. In *Aspartic Proteinases*; Springer: Dordrecht, The Netherlands, 1995; pp 279−283.

(18) Travis, J.; Fritz, H. Potential problems in designing elastase inhibitors for therapy. *Am. Rev. Respir. Dis.* **1991**, *143*, 1412−1415.
(19) Prin-Mathieu, C.; Le Roux, Y.; Faure, G.; Laurent, F.; Béné, M.; Moussaoui, F. Enzymatic activities of bovine peripheral blood leukocytes and milk polymorphonuclear neutrophils during intramammary inflammation caused by lipopolysaccharide. *Clin. Diagn. Lab. Immunol.* **2002**, *9*, 812−817.
(20) Kelly, A.; O'Flaherty, F.; Fox, P. Indigenous proteolytic enzymes in milk: a brief overview of the present state of knowledge. *Int. Dairy J.* **2006**, *16*, 563−572.
(21) Hayes, M.; Stanton, C.; Fitzgerald, G. F.; Ross, R. P. Putting microbes to work: dairy fermentation, cell factories and bioactive peptides. Part II: Bioactive peptide functions. *Biotechnol. J.* **2007**, *2*, 435−449.
(22) Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J.-C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. Protein identification and analysis tools in the ExPASy server. In *2-D Proteome Analysis Protocols*; Springer: Dordrecht, The Netherlands, 1999; pp 531−552.
(23) Christensen, B.; Schack, L.; Kläning, E.; Sørensen, E. S. Osteopontin is cleaved at multiple sites close to its integrin-binding motifs in milk and is a novel substrate for plasmin and cathepsin D. *J. Biol. Chem.* **2010**, *285*, 7929−7937.